SAND2022-1647C

Assessing IO500 Benchmark for Performance Engineering Guidance

Radita Liem*, Jay Lofstead+, Julian Kunkel

*Chair for High Performance Computing, IT Center, RWTH Aachen University
+ Sandia National Laboratories

Motivation

Evaluating I/O performance is notoriously difficult to do due to various intertwined variables. To answer this challenge, there are several benchmarks available out there with IO500 benchmark[1] is the current de-facto leader of the benchmark.

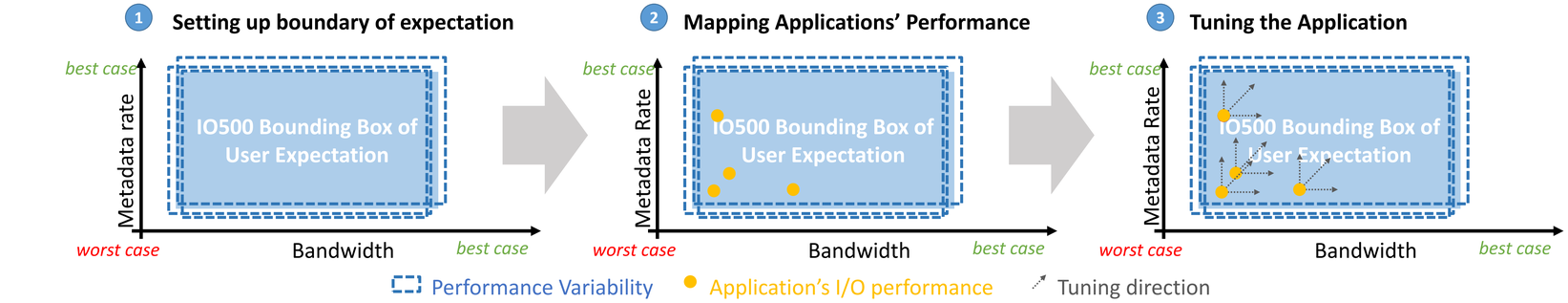
We use IO500 benchmark based workflow that we proposed in our previous work[2][3] and the public data available in the IO500 website to showcase how the real-world systems and various applications interact with the benchmark. This work aims to provide more insights for IO500 users that can help them understand benchmark numbers and use it for performance improvements

IO500 Workflow

In our proposed workflow, we use the IO500 benchmark result from the **'easy'** and **'hard'** case to form bounding-box of I/O performance expectation.

The 'easy' case is a free to tune parameters represents the best case scenario for bandwidth and metadata. The 'hard' case has limited options to tune for worst case scenario.

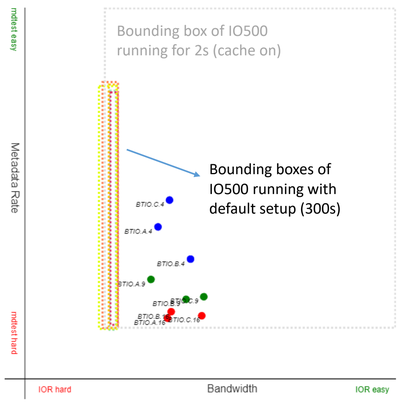
By placing application's performance information into this bounding box. Users can get information on the state of their application and plan the tuning strategy accordingly.



Performance Mapping

IO500 with MPI-IO API Result

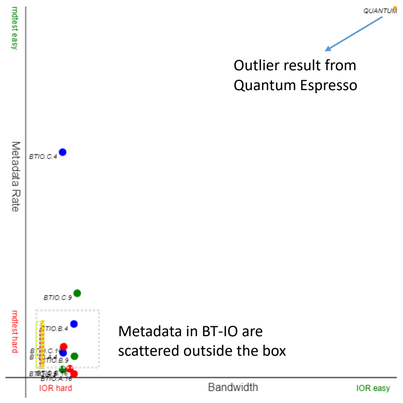
Application's bandwidth for MPI-IO API falls outside of the IO500 box because of the page caching. When we enable the cache, the application performance falls within the box



Benchmark	Bandwidth (GiB/s)	Metadata Rate (KIOPS)
BTIO.A.4	0.68	21.67
BTIO.A.9	0.64	14.18
BTIO.A.16	0.73	8.66
BTIO.B.4	0.85	17.08
BTIO.B.9	0.82	11.35
BTIO.B.16	0.75	9.58
BTIO.C.4	0.74	25.50
BTIO.C.9	0.91	11.71
BTIO.C.16	0.90	9.00

IO500 with POSIX Result

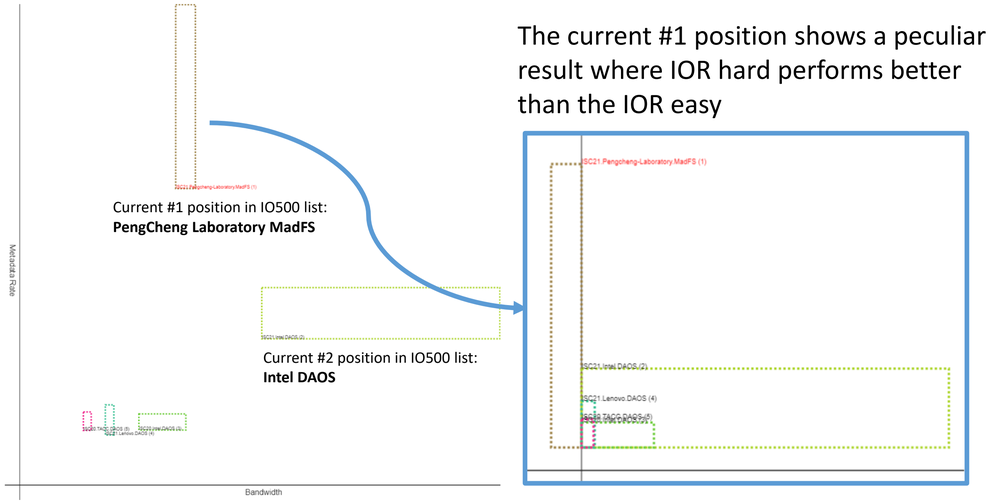
Quantum Espresso performance falls far outside the bounding box of user expectation. For BT-IO benchmark, the bandwidth is within the box with scattered metadata.



Benchmark	Bandwidth (GiB/s)	Metadata Rate (KIOPS)
BTIO.A.4	0.99	18.42
BTIO.A.9	0.97	5.99
BTIO.A.16	1.27	2.69
BTIO.B.4	1.27	39.90
BTIO.B.9	1.28	15.84
BTIO.B.16	1.16	6.51
BTIO.C.4	0.97	167.96
BTIO.C.9	1.36	62.81
BTIO.C.16	1.00	22.99
QUANTUM ESPRESSO.24	9.77	276.59

Observation on IO500 List

Within the top five of IO500 ten nodes challenge, there are two outlier system that performs so much better than the rest of the submission.



Conclusion and Future Outlook

- There are needs to have other I/O performance data from various applications to give meanings on the IO500 result → can setup a better configuration on the IO500 benchmark
- Changes in the system designs made the current hard case might be irrelevant → however, we need to assess whether the system still works well with the traditional workload
- Page caching data needs to be addressed in the IO500 benchmark itself → perhaps there are mode and guideline for enable caching so users can have a complete view on multiple scenarios

[1] J. Kunkel, "Virtual Institute for I/O," *IO-500*, 30-Oct-2020. [Online]. Available: https://www.vi4io.org/std/io500/start. [Accessed: 02-Mar-2021].
[2] D. Povaliaiev, R. Liem, J. Loftstead, C. Terboven. An IO500-based Workflow For User-centric I/O Performance Management. Poster presented at: ISC2021.
[3] R. Liem, D. Povaliaiev, J. Lofstead, J. Kunkel and C. Terboven, "User-Centric System Fault Identification Using IO500 Benchmark," in 2021 IEEE/ACM Sixth International Parallel Data Systems Workshop (PDSW), St. Louis, MO, USA, 2021 pp. 35-40. doi: 10.1109/PDSW54622.2021.00011
[4] M. Rásó-Barnett, "Lustre and IO-500: Experiences with the Cambridge Data Accelerator", 2019. [Online]. https://www.eofs.eu/_media/events/lad19/03_matt_raso-barnett-io500-cambridge.pdf. [Accessed: 02-Mar-2021]
[5] A. Dilger, "IO500 | A storage Benchmark for HPC", 2019. [Online]. Available:https://wiki.lustre.org/images/9/92/LUG2019-IO500_Storage_Benchmark_for_HPC-Dilger.pdf. [Accessed: 02-Mar-2021]
[6] BeeGFS – The Leading Parallel Cluster File System," *BeeGFS*, 01-Mar-2021. [Online]. Available: https://www.beegfs.io/c/. [Accessed: 08-Mar-2021].

liem@itc.rwth-aachen.de
https://www.i12.rwth-aachen.de/